

# Klasifikasi Tingkat Partisipasi Angkatan Kerja Kota Makassar Menggunakan Metode CART

Adnan Sauddin

Universitas Islam Negeri Alauddin Makassar, [adnan.sauddin@uin-alauddin.ac.id](mailto:adnan.sauddin@uin-alauddin.ac.id)

Wahidah Alwi

Universitas Islam Negeri Alauddin Makassar, [adnan.sauddin@uin-alauddin.ac.id](mailto:adnan.sauddin@uin-alauddin.ac.id)

A.Ningsih AN

Universitas Islam Negeri Alauddin Makassar, [adnan.sauddin@uin-alauddin.ac.id](mailto:adnan.sauddin@uin-alauddin.ac.id)

**ABSTRAK.** Penelitian ini membahas tentang kondisi angkatan kerja di kota Makassar yang mengalami peningkatan setiap tahunnya. Metode yang digunakan adalah metode CART (*Classification And Regression Tree*). Adapun hasil dari penelitian ini diperoleh bahwa metode CART dengan tingkat ketepatan klasifikasi 69,68% mengklasifikasi angkatan kerja menjadi lima .

**Kata Kunci:** kondisi iklim, curah hujan, time series, SARIMA

## 1. PENDAHULUAN

Makassar Makassar sebagai kota metropolitan terbesar di Indonesia timur yang merupakan kota dengan jumlah penduduk terbesar di Sulawesi Selatan. Berdasarkan data dari Badan Pusat Statistik kota Makassar tercatat jumlah penduduk kota Makassar pada Tahun 2016 sebesar 1.469.601 jiwa, dan pada Tahun 2017 meningkat menjadi 1.489.011 jiwa. Dari data tersebut dapat dilihat pertumbuhan penduduk di kota Makassar mengalami peningkatan sebesar 1,32% dari tahun 2016 sampai 2017. Peningkatan jumlah penduduk ini menjadi salah satu penyebab terjadinya pembengkakan jumlah angkatan kerja di kota Makassar. Hal ini disebabkan banyaknya penduduk yang mencari pekerjaan.

Tingkat partisipasi angkatan kerja di kota Makassar pada Tahun 2016 sampai 2017 mengalami peningkatan dimana Tingkat Partisipasi Angkatan Kerja meningkat dari 55,20 % pada Tahun 2016 meningkat menjadi 55,23 % di Tahun 2017. Tingkat pendidikan angkatan kerja di kota Makassar relatif baik dilihat dari total angkatan kerja, sekitar 27,17% merupakan lulusan diploma dan Universitas, 37,83% tamatan SMA/SMK, 15,02% tamatan SMP, dan selebihnya 19,98% tamatan SD, tidak tamat SD, dan tidak pernah sekolah. jika tingkat pendidikan angkatan kerja dilihat dari total jumlah penduduk yang bekerja sekitar 27,24%

merupakan lulusan diploma dan Universitas, 36,19 % tamatan SMA/SMK, 15,04 % tamatan SMP, dan selebihnya 21,53% tamatan SD, tidak tamat SD, dan tidak pernah sekolah.

Tingkat Partisipasi Angkatan Kerja di kota Makassar jika dipandang dari jenis kelamin pada tahun jumlah 2016 jumlah Tingkat Partisipasi Angkatan Kerja laki-laki sebesar 70,22% sedangkan Tingkat Partisipasi Angkatan Kerja perempuan sebesar 40,84%. Pada tahun 2017 jumlah Tingkat Partisipasi Angkatan Kerja laki-laki mengalami peningkatan menjadi 72,90% sedangkan Tingkat Partisipasi Angkatan Kerja perempuan mengalami penurunan menjadi 38,28%. Dari data tersebut dapat dilihat bahwa Tingkat Partisipasi Angkatan kerja Laki-laki masih mendominasi pasar tenaga kerja.

Berdasarkan uraian di atas dapat dilihat bahwa penyerapan angkatan kerja dapat dipengaruhi oleh faktor usia produktif, jenis kelamin, tingkat pendidikan, status pernikahan, dan penyerapan angkatan kerja juga dipengaruhi oleh pengalaman pelatihan kerja seperti pelatihan keterampilan atau kursus dimana dengan memiliki pengalaman kerja didukung tingkat pendidikan yang tinggi, maka tenaga kerja akan mempunyai lebih banyak kesempatan untuk mendapatkan pekerjaan. Sehingga untuk mengetahui besar Tingkat Partisipasi Angkatan Kerja di kota Makassar berdasarkan faktor-faktor tersebut maka Tingkat Partisipasi Angkatan Kerja di kota Makassar akan dikelompokkan menjadi angkatan kerja yang bekerja dan tidak bekerja didasarkan pada faktor yang mempengaruhinya.

Terdapat beberapa metode yang sering digunakan untuk mengklasifikasi atau mengelompokkan data seperti CART, CHAID, QUEST, dan C4.5. Metode-metode ini telah banyak digunakan sebelumnya dalam beberapa penelitian dan diperoleh bahwa metode CART

memiliki ketepatan hasil klasifikasi yang lebih baik dibandingkan metode yang lain. Metode *Classification and Regression Tree* (CART) telah lama digunakan untuk memperoleh pohon keputusan dengan cara membentuk pohon klasifikasi. Metode CART dapat digunakan untuk mendapatkan suatu kelompok data yang akurat sebagai penciri dari pengklasifikasian. Selain itu metode CART juga dapat digunakan pada data yang variabelnya dapat terdiri dari data kontinu dan kategorik, nominal, maupun ordinal dengan jumlah data yang besar menggunakan prosedur pemilah biner sehingga dapat dengan mudah melihat pola data yang telah diteliti. Oleh sebab itu, dalam penelitian ini akan diterapkan metode CART dalam mengklasifikasi angkatan kerja di kota Makassar karena variabel yang akan digunakan dalam penelitian ini yaitu variabel usia produktif, jenis kelamin, tingkat pendidikan, status pernikahan, dan pengalaman pelatihan kerja jenis datanya kategorik dan data akan diklasifikasi menjadi dua kelompok yaitu angkatan kerja yang bekerja dan tidak bekerja.

## 2. TINJAUAN PUSTAKA

### Data Mining

Data mining adalah proses seleksi, eksplorasi, dan pemodelan yang menganalisis kumpulan data yang besar digunakan dalam menemukan pola atau kecenderungan dari data yang biasanya tidak disadari keberadaannya. Data mining adalah bagian dari KDD (*Knowledge Discovery in Database*) yaitu proses mengubah data mentah menjadi data yang dapat digunakan sebagai tambahan informasi atau pengetahuan yang baru yang berguna dan dapat dimengerti. Data mining merupakan proses pemilihan atau “menambang” pengetahuan dari sekumpulan data yang berjumlah banyak atau besar. Proses data mining dapat dilakukan dalam beberapa langkah yaitu pembersihan data, integrasi data, seleksi data, transformasi data.

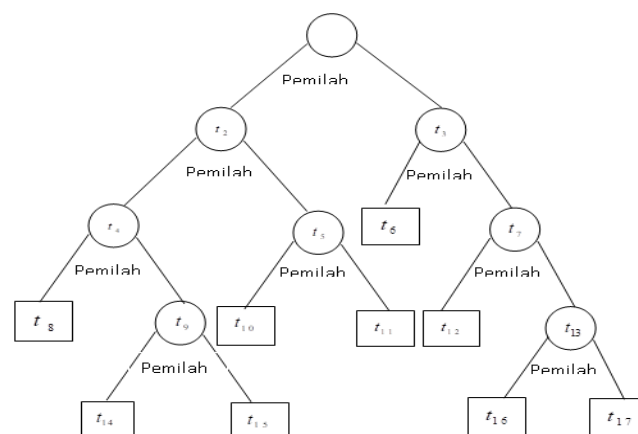
### Decision Tree

*Decision Trees* merupakan salah satu teknik yang digunakan dalam melakukan klasifikasi terhadap sekumpulan beberapa objek. Teknik ini terdiri dari sekumpulan *decision node* yang pada tahap ini mengevaluasi semua atribut, dihubungkan oleh cabang, bergerak kebawah dari *root node* sampai berakhir di *leaf node*. Tiap cabang dapat

masuk ke *decision node* yang lain ataupun pada *leaf node*. *Decision Trees* adalah flowchart seperti struktur pohon yang berusaha menemukan atribut yang tepat untuk menyelesaikan dan menentukan kelas. Beberapa algoritma pohon keputusan dapat menghasilkan pohon keputusan biner dimana setiap node internal dalam pohon ini bercabang tepat menjadi dua node yang lain. Sedangkan algoritma pohon keputusan yang lain juga dapat menghasilkan pohon keputusan non biner dimana setiap node internal dalam pohon ini bercabang menjadi dua atau lebih node yang lain

### Metode CART

*Classification And Regression Trees* merupakan metode klasifikasi menggunakan data historis untuk membangun suatu pohon keputusan. CART merupakan suatu analisis nonparametrik yang digunakan untuk menyajikan keputusan yang berbentuk pohon biner. Tujuan utama CART adalah untuk mendapatkan suatu kelompok data yang akurat sebagai penciri dari suatu pengklasifikasian. Pada setiap terminal node diberikan label kelas atau nilai yang diramalkan, sehingga menghasilkan struktur pohon yang dapat ditafsirkan sebagai pohon keputusan.



Gambar 2.2 Pohon Klasifikasi CART

Keterangan :

1. *Root node* digambarkan dengan lingkaran merupakan *nonterminal node* paling awal tempat *learning sample* yang dimiliki.
2. *Branch* digambarkan dengan 2 garis lurus merupakan cabang dari *root node*.

*Branch* merupakan tempat pemecahan masing-masing *nonterminal node*.

3. *Nonterminal nodes* digambarkan dengan lingkaran merupakan subset atau himpunan bagian dari *nonterminal node* di atasnya yang memenuhi kriteria pemecahan tertentu.
4. *Terminal Nodes*, digambarkan dengan persegi merupakan node tempat memprediksi kan sebuah objek pada kelas tertentu (*class labeled*).

### Struktur Pohon Klasifikasi CART

Proses analisis dalam CART digambarkan dalam bentuk atau struktur yang menyerupai sebuah pohon *biner*. *Biner* dalam pohon klasifikasi ini berarti setiap pemecahan *parent node* menghasilkan 2 *child nodes*. Simpul awal yang disebut *parent node* dinotasikan dengan  $t_1$ . Pohon klasifikasi CART dapat dilihat pada Gambar 2.1.

### Algoritma CART

Proses analisis dari metode CART dalam membuat pohon klasifikasi dikenal dengan istilah *Binary Recursive Partitioning*. Istilah *binary* mengacu bahwa setiap *parent node* akan terbagi menjadi dua kelompok *child node*. Sedangkan istilah *recursive* mengacu pada fakta bahwa proses pemecahan dilakukan berulang kali. Dengan demikian *parent node* yang menghasilkan dua *child nodes*, selanjutnya masing-masing *child nodes* tersebut mungkin akan melakukan pemecahan dan menghasilkan *child nodes* tambahan. Istilah *Partitioning* mengacu pada fakta bahwa dataset dibagi menjadi beberapa bagian atau dipartisi. Data yang digunakan dalam metode CART adalah data training dan data testing. Data training adalah data yang siap untuk dimining yang telah melewati data preprocessing sedangkan data testing adalah data yang digunakan untuk menguji rule klasifikasi yang diperoleh dari data training. sehingga Algoritma CART melalui tiga tahapan sebagai berikut:

1. Pembentukan pohon klasifikasi

Dalam pembentukan pohon klasifikasi terdiri dari tiga proses yaitu pemilihan atribut, penentuan simpul terminal, dan penandaan label kelas seperti berikut:

#### a. Pemilihan pemilah

Data yang digunakan merupakan sampel data *learning* yang masih bersifat heterogen. Sampel tersebut akan dipilih berdasarkan aturan pemilihan dan kriteria *goodness-of-split* dan pemilihan pemilah bergantung pada jenis variabel terikatnya. Sebelum melakukan pemilihan pemilah dengan menggunakan indeks gini, akan lebih baik mencari *gain information* dari setiap simpul dengan menggunakan rumus seperti berikut:

$$GI(t) = - \sum_{j=1}^n P(j|t) \log_2 P(j|t) \quad (2.1)$$

Keterangan :

$GI(t)$  = Gain Information pada simpul  $t$   
 $P(j|t)$  = Proporsi kelas  $j$  pada simpul  $t$  dimana  $j=1,2,3, \dots, n$  dengan  $P(j|t) = \frac{n_j(t)}{n(t)}$ .

$n_j(t)$  = Banyak pengamatan kelas  $j$  pada simpul  $t$

$n(t)$  = Banyak pengamatan pada simpul  $t$

Metode pemilihan pemilah pada metode CART menggunakan indeks gini  $i(t)$  yang merupakan pengukuran tingkat keragaman suatu kelas dari suatu simpul tertentu dalam klasifikasi sehingga mampu membantu dalam menemukan fungsi pemilah yang optimal. Indeks Gini adalah pengukuran keragaman antara probabilitas nilai-nilai atribut target. Fungsi Indeks Gini sebagai berikut:

$$i(t) = 1 - \sum_{j=1}^n P^2(j|t) \quad (2.2)$$

Keterangan :

$i(t)$  = Indeks gini

$P(j|t)$  = Proporsi kelas  $j$  pada simpul  $t$  dimana  $j=1,2,3, \dots, n$

dengan  $P(j|t) = \frac{n_j(t)}{n(t)}$ .

$n_j(t)$  = Banyak pengamatan kelas  $j$  pada simpul  $t$

$n(t)$  = Banyak pengamatan pada simpul  $t$

Atribut yang terpilih akan membentuk suatu himpunan kelas yang disebut simpul atau node. simpul akan selalu mengalami pemilihan secara rekursif sampai diperoleh terminal nodes. Langkah selanjutnya adalah menentukan kriteria *goodness-of-split* yang merupakan suatu evaluasi pemilihan oleh pemilah  $s$  pada  $t$  yang disebut juga sebagai penurunan keheterogenan dengan rumus sebagai berikut:

$$\phi(s, t) = i(t) - P_L i(t_L) - P_R i(t_R) \quad (2.3)$$

Dimana:

$t_R$  = Simpul kanan atau *right node*

$t_L$  = Simpul kiri atau *left node*

$P_R$  = Proporsi banyaknya objek yang pada

$t_R$

$P_L$  = Proporsi banyaknya objek yang pada  $t_L$

Probabilitas banyaknya objek yang masuk pada  $t_R$  dan  $t_L$  dapat dihitung dengan:

$$P_L = \frac{\text{calon simpul kiri}}{\text{data latihan}} \quad (2.4)$$

$$P_R = \frac{\text{calon simpul kanan}}{\text{data latihan}} \quad (2.5)$$

Pengembangan pohon dilakukan dengan mencari semua kemungkinan pemilah pada simpul  $t$ , yang kemudian akan ditemukan pemilah  $s^*$  yang memberikan nilai penurunan keheterogenan tertinggi yaitu:

$$\phi(s^*, t) = \max_{s \in S} \phi(s, t) \quad (2.6)$$

#### b. Penentuan Simpul Terminal

Suatu simpul  $t$  akan menjadi simpul terminal atau tidak apabila pada simpul  $t$  tidak terdapat penurunan keheterogenan secara berarti berdasarkan kriteria *goodness-of-split* atau hanya terdapat satu pengamatan pada tiap simpul anak atau jumlah kasus minimum dalam suatu terminal akhir pada pengamatan berjumlah kurang atau sama dengan 5 ( $n \leq 5$ ). Selain itu, proses pembentukan pohon juga akan berhenti apabila sudah mencapai batasan jumlah level yang telah ditentukan atau tingkat kedalaman dalam pohon maksimal. Serta proses pemecahan pohon pada node  $t$  menjadi  $t_R$  dan  $t_L$  berlaku:

$$R(t) > R(t_R) + R(t_L) \quad (2.7)$$

Dimana :

$R(t)$  = Kesalahan pada node  $t$ .

$R(t_L)$  = Kesalahan pada simpul kiri node  $t$ .

$R(t_R)$  = Kesalahan pada simpul kanan node  $t$ .

#### c. Penandaan Label Kelas

Penandaan label kelas adalah proses pengidentifikasian tiap nodes pada suatu kelas tertentu. Penandaan kelas dilakukan pada *terminal nodes*, *nonterminal nodes*, dan *root nodes*. Akan tetapi, penandaan label paling dibutuhkan pada terminal nodes atau simpul terminal karena simpul ini penting digunakan untuk memprediksi suatu objek pada kelas tertentu yang berada pada simpul terminal ini. Penandaan label kelas pada simpul terminal dilakukan berdasarkan aturan jumlah terbanyak yaitu jika:

$$P(j|t) = \max_j \frac{N_j(t)}{N(t)} \quad (2.8)$$

Dimana :

$P(j|t)$  = Proporsi kelas  $j$  pada simpul  $t$

$N_j(t)$  = Jumlah Pengamatan kelas pada simpul  $t$

$N(t)$  = Jumlah Pengamatan simpul  $t$

#### 2. Pemangkasan Pohon Klasifikasi

Pohon yang terbentuk dengan aturan pemilah dan kriteria *goodness of split* berukuran sangat besar karena pemberhentian pohon berdasarkan banyaknya pengamatan pada simpul terminal atau besarnya tingkat kehomogenan. Untuk mendapatkan ukuran pohon yang layak dapat dilakukan dengan melakukan pemangkasan pohon dengan berdasarkan ukuran *cost complexity minimum* menggunakan rumus sebagai berikut:

$$g_m(t) = \frac{R(t) - R(T_k)}{|T_k| - 1} \quad (2.9)$$

Dimana :

$g_m(t)$  = Complexity parameter

$R(t)$  = Kesalahan pengklasifikasian pada node  $t$ .

$T_k$  = Subtree ke- $k$ , dengan  $k=1,2,...,n$

$R(T_k)$  = Kesalahan pengklasifikasian pada pohon  $T_k$  dengan  $R(T_k) = \sum_m R(t_{m_k})$

$R(t_{m_k})$  = Kesalahan pengklasifikasian pada simpul  $t$  ke  $m$  di pohon ke- $k$

$\tilde{T}_k$  = Simpul terminal pada  $T_k$

$\left| \tilde{T}_k \right|$  = Banyaknya simpul pada pohon  $T_k$

Cabang yang dipangkas adalah cabang yang memiliki nilai  $g_m(t)$  terkecil, yaitu:

$$g_m(t_m) = \min_{t \in T_k} g_m(t) \quad (2.10)$$

Kesalahan pengklasifikasian pada node  $t$  diperoleh dengan menggunakan rumus sebagai berikut:

$$R(t) = r(t) \cdot P(t) \quad (2.11)$$

Dimana:

$r(t)$  = Probabilitas atau terjadinya pengklasifikasian pada node dengan  $r(t) = 1 - \max(j|t)$

$P(t)$  = Proporsi banyaknya objek simpul  $t$

### 3. Penentuan Pohon Klasifikasi Optimal

Ukuran pohon yang besar akan menyebabkan nilai kompleksitas yang tinggi karena struktur data yang digambarkan cenderung kompleks, sehingga perlu dilakukan pemilihan pohon optimal yang berukuran sederhana tetapi dapat memberikan nilai penduga pengganti cukup kecil. Penduga yang sering digunakan adalah penduga validasi silang lipat  $V$  (*cross validation V-fold estimate*).

Penduga validasi silang lipat  $V$  digunakan pada data yang tidak cukup besar yaitu  $< 3000$  (kurang dari 3000 data). Dalam *cross validation V-fold estimate*, pengamatan dalam  $L$  dibagi secara random menjadi  $V$  bagian yang saling lepas dengan ukuran kurang lebih sama besar untuk setiap kelas. Pohon  $T^v$  dibentuk dari sampel *learnig* ke- $v$  dimana  $v = 1, 2, 3, \dots, V$ . Penduga sampel uji untuk  $R(T_i^{(v)})$  sebagai berikut:

$$R(T_i^{(v)}) = \frac{1}{N_v} \sum_v X(d^{(v)}) \quad (2.12)$$

Dimana:

$X(d^{(v)})$  = Hasil pengklasifikasian

$N_v$  = Jumlah pengamatan dalam  $L_v$

Pohon klasifikasi optimal yang dipilih adalah pohon klasifikasi yang memiliki *cross validation* yang paling kecil yang berarti pohon yang mempunyai tingkat kesalahan pengklasifikasian paling kecil yang memenuhi persamaan:

$$TKK = 1 - \frac{n_{11} + n_{22}}{N} \quad (2.13)$$

Dimana:

TKK = Tingkat Kesalahan Klasifikasi

$n_{11}$  = Jumlah observasi dari kelas yang tepat diprediksi kelas 1

$n_{22}$  = Jumlah Observasi dari kelas 2 yang tepat diprediksi kelas 2

### Ukuran Ketepatan Klasifikasi

Ukuran ketepatan klasifikasi digunakan untuk mengetahui apakah data yang diklasifikasikan dengan benar atau tidak. beberapa cara yang biasa digunakan dalam mengukur ketepatan klasifikasi adalah melalui perhitungan, *Sensitivity*, *specivity*, dan akurasi:

1. *Sensitivity* digunakan untuk mengukur ketepatan klasifikasi pada kelas  $I$ ,
2. *specivity* digunakan untuk mengukur ketepatan klasifikasi pada sampel kelas  $j$ .
3. Akurasi digunakan untuk mengukur tingkat ketepatan klasifikasi

Tabel yang digunakan dalam menghitung ketepatan klasifikasi ditujukan sebagai berikut:

Tabel 2.1 Ukuran Ketepatan Klasifikasi

Observasi Y	Prediksi Y		Total
	1	2	
1	$n_{11}$	$n_{12}$	$N_1$
2	$n_{21}$	$n_{22}$	$N_2$
Total	$N_1$	$N_2$	$N$

Dengan menggunakan tabel diatas nilai *Sensitivity*, *specivity*, *akurasi*, dan tingkat kesalahan klasifikasi dapat dihitung dengan menggunakan rumus sebagai berikut:

$$Sensitivity = \frac{n_{11}}{N_1} \quad (2.14)$$

$$Specificity = \frac{n_{22}}{N_2} \quad (2.15)$$

$$Akurasi = \frac{n_{11} + n_{22}}{N} \quad (2.16)$$

Dimana:

$n_{11}$  = Jumlah observasi dari kelas 1 yang tepat diprediksi sebagai kelas 1

$n_{12}$  = Jumlah Observasi dari kelas 1 yang tepat diprediksi sebagai kelas 2

$n_{21}$  = Jumlah Observasi dari kelas 2 yang tepat diprediksi sebagai kelas 1

$n_{22}$  = Jumlah Observasi dari kelas 2 yang tepat diprediksi sebagai kelas 2

$N_1$  = Jumlah Observasi dari kelas 1

$N_2$  = Jumlah observasi dari kelas 2

### 3. METODOLOGI

Jenis penelitian yang digunakan adalah penelitian terapan dengan Jenis data dalam penelitian ini adalah data sekunder dan sumber data yaitu data yang diambil langsung dari Badan pusat Statistik (BPS) Provinsi Sulawesi Selatan untuk periode 2018. Penelitian ini dilakukan di Badan Pusat Statistik (BPS) Provinsi Sulawesi Selatan. Penelitian ini dilakukan mulai dari Januari-Juli 2019. Variabel yang digunakan dalam penelitian terdiri dari :

#### 1. Variabel Bebas (Independen)

Variabel bebas dalam penelitian ini adalah sebagai berikut :

Variabel	Nama Variabel	Skala	Kategori
$X_1$	Jenis Kelamin	Nominal	1 : Laki-laki 2 : Perempuan
$X_2$	Tingkat Pendidikan	Ordinal	1: Rendah ( $\leq SD$ ) 2 : Sedang ( $SMP - SMA$ ) 3 : Tinggi ( $Sarjana$ )
$X_3$	Status Pernikahan	Nominal	1: Belum Menikah 2 : Menikah
$X_4$	Pengalaman Pelatihan Kerja	Nominal	1 : Pernah 2: Tidak Pernah

#### 2. Variabel Terikat (Dependen)

Variabel terikat dalam penelitian ini berskala biner yaitu :

$$Y = \begin{cases} 1: & \text{Angkatan kerja yang bekerja} \\ 0: & \text{Angkatan kerja yang tidak bekerja} \end{cases}$$

### Prosedur Analisis

Analisis data yang digunakan pada penelitian ini sebagai berikut:

1. Membagi data menjadi dua bagian yaitu data *learning* dan data *testing*.
2. Membentuk pohon klasifikasi maksimal dengan langkah-langkah sebagai berikut:
  - a. Melakukan pemilihan pemilah berdasarkan aturan pemilihan indeks gini yang selanjutnya dievaluasi menggunakan kriteria *goodness of split*.
  - b. Menentukan simpul terminal atau *terminal node*.
  - c. Melakukan penandaan label kelas pada terminal node berdasarkan aturan jumlah terbanyak dari tiap kelas yang ada pada variabel respon.
3. Pemangkasan pohon klasifikasi berdasarkan kriteria *complexity parameter minimum*.
4. Memilih pohon klasifikasi optimal.
5. Menghitung ketepatan klasifikasi pohon dengan menggunakan *sensitivity*, *specificity*, dan akurasi.
6. Menghitung jumlah tingkat partisipasi angkatan kerja yang bekerja dan tidak bekerja berdasarkan pohon klasifikasi yang terbentuk.

### 4. PEMBAHASAN

#### Hasil Penelitian

##### Deskripsi Data

Data yang digunakan dalam penelitian ini adalah data hasil Survei Angkatan Kerja Nasional Tahun 2018 dikota Makassar yang diperoleh dari Badan Pusat Statistik Provinsi Sulawesi Selatan. Tabel 4.1 berikut ini menampilkan rangkuman jumlah angkatan kerja berdasarkan jenis kelamin, yaitu:



Tabel 4.1 Jumlah angkatan kerja berdasarkan jenis kelamin

		Status Angkatan Kerja				Total
		Angkatan kerja yang bekerja		Angkatan kerja yang tidak bekerja		
		N	%	N	%	
Jenis Kelamin	Laki-laki	504	64,86%	273	35,14%	777
	Perempuan	273	35,32%	500	64,68%	773
Total						1550

Berdasarkan Tabel 4.1 di atas dapat dijelaskan bahwa jumlah angkatan kerja laki-laki yang bekerja adalah sebanyak 504 (64,86%, n=1550) sedangkan jumlah angkatan kerja laki-laki yang tidak bekerja adalah sebanyak 273 (35,14%, n=1550). Untuk jumlah angkatan kerja perempuan yang bekerja adalah sebanyak 273 (35,32%, n=1550) sedangkan jumlah angkatan kerja perempuan yang tidak bekerja adalah sebanyak 500 (64,68%, n=1550). Tabel 4.2 berikut ini menampilkan rangkuman jumlah angkatan kerja berdasarkan tingkat pendidikan, yaitu:

Tabel 4.2 Jumlah angkatan kerja berdasarkan tingkat pendidikan

		Status Angkatan Kerja				Total
		Angkatan kerja yang bekerja		Angkatan kerja yang tidak bekerja		
		N	%	N	%	
Tingkat Pendidikan	Rendah	186	53,45%	162	46,55%	348
	Sedang	379	44,02%	482	55,98%	861
	Tinggi	212	62,17%	129	37,83%	341
Total						1550

Berdasarkan Tabel 4.2 di atas dapat dijelaskan bahwa jumlah angkatan kerja yang bekerja pada tingkat pendidikan yang tergolong rendah sebanyak 186 (53,45%, n=1550) sedangkan jumlah angkatan kerja yang tidak bekerja sebanyak 162 (46,55%, n=1550). Angkatan kerja yang bekerja pada tingkat pendidikan yang tergolong sedang sebanyak 379 (44,02%, n=1550) sedangkan jumlah angkatan kerja yang tidak bekerja sebanyak 482 (55,98%, n=1550). Angkatan kerja yang bekerja pada tingkat pendidikan yang tergolong tinggi sebanyak 212 (62,17%, n=1550) sedangkan jumlah angkatan kerja yang tidak bekerja sebanyak 129 (37,83%, n=1550). Tabel 4.3 berikut ini menampilkan rangkuman jumlah

angkatan kerja berdasarkan status pernikahan, yaitu:

Tabel 4.3 Jumlah angkatan kerja berdasarkan status pernikahan

		Status Angkatan Kerja				Total
		Angkatan kerja yang bekerja		Angkatan kerja yang tidak bekerja		
		N	%	N	%	
Status Pernikahan	Belum Menikah	236	39,73%	358	60,27%	594
	Menikah	541	56,59%	415	43,41%	956
Total						1550

Berdasarkan Tabel 4.3 di atas bahwa jumlah angkatan kerja yang bekerja dengan status pernikahan belum menikah adalah sebanyak 236 (39,73%, n=1550) sedangkan jumlah angkatan kerja yang tidak bekerja adalah sebanyak 358 (60,27%, n=1550). Untuk jumlah angkatan kerja yang bekerja dengan status pernikahan menikah adalah sebanyak 541 (56,59%, n=1550) sedangkan jumlah angkatan kerja yang tidak bekerja adalah sebanyak 415 (43,41%, n=1550). Tabel 4.4 berikut ini menampilkan rangkuman jumlah angkatan kerja berdasarkan pengalaman pelatihan kerja, yaitu:

Tabel 4.4 Jumlah angkatan kerja berdasarkan pengalaman pelatihan kerja

		Status Angkatan Kerja				Total
		Angkatan kerja yang bekerja		Angkatan kerja yang tidak bekerja		
		N	%	N	%	
Pengalaman	Pernah	125	60,39%	82	39,61%	207
Pelatihan Kerja	Tidak	652	48,55%	691	51,45%	1343
	Pernah					
Total						1550

Berdasarkan Tabel 4.4 di atas bahwa jumlah angkatan kerja yang bekerja dengan pengalaman pelatihan kerja pernah adalah sebanyak 125 (60,39%, n=1550) sedangkan jumlah angkatan kerja yang tidak bekerja adalah 82 (39,61%, n=1550). Untuk jumlah angkatan kerja yang bekerja dengan pengalaman pelatihan kerja tidak pernah adalah sebanyak 652 (48,55%, n=1550) sedangkan jumlah angkatan kerja yang tidak bekerja adalah 691 (51,45%, n=1550). Gambar 4.1 berikut ini menampilkan rangkuman jumlah angkatan kerja yang bekerja

dan tidak bekerja berdasarkan jenis kelamin dan tingkat pendidikan, yaitu:

		Jenis Kelamin		0	1
		Laki-Laki	Perempuan		
Tingkat Pendidikan	Rendah	147	172	155	164
	Sedang	440	350	402	388
	Tinggi	109	177	139	147
Grand Total		696	699	696	1395

Gambar 4.1 Data Cube Berdasarkan Jenis Kelamin dan Tingkat Pendidikan

Berdasarkan Gambar 4.1 di atas diperoleh bahwa jumlah angkatan kerja yang bekerja dengan jenis kelamin laki-laki dan perempuan masing-masing adalah 456 dan 240 sedangkan jumlah angkatan kerja yang tidak bekerja masing-masing adalah 456 dan 243. Untuk jumlah angkatan kerja yang bekerja dengan tingkat pendidikan rendah, sedang dan tinggi masing-masing adalah 172, 359 dan 177 sedangkan jumlah angkatan kerja yang tidak bekerja masing-masing adalah 147, 440, dan 109. Untuk jumlah angkatan kerja jenis kelamin laki-laki dengan tingkat pendidikan rendah, sedang, dan tinggi masing-masing adalah 164, 388, dan 147 sedangkan untuk jumlah angkatan kerja jenis kelamin perempuan dengan tingkat pendidikan rendah, sedang, dan tinggi masing-masing adalah 155, 402, dan 139. Gambar 4.2 berikut ini menampilkan rangkuman jumlah angkatan kerja yang bekerja dan tidak bekerja berdasarkan jenis kelamin dan status pernikahan, yaitu:

		Jenis Kelamin		0	1
		Laki-Laki	Perempuan		
Status Pernikahan	Belum Menikah	324	213	303	234
	Menikah	372	486	393	465
	Grand Total	696	699	696	1395

Gambar 4.2 Data cube berdasarkan jenis kelamin dan status pernikahan

Berdasarkan Gambar 4.2 di atas diperoleh bahwa jumlah angkatan kerja yang bekerja dengan jenis kelamin laki-laki dan perempuan masing-masing adalah 456 dan 240 sedangkan jumlah angkatan kerja yang tidak bekerja masing-masing adalah 456 dan 243. Untuk jumlah angkatan kerja yang bekerja dengan status pernikahan belum menikah masing-masing adalah 324 dan 372 sedangkan jumlah angkatan kerja yang tidak bekerja masing-masing adalah 213 dan 486.

Untuk jumlah angkatan kerja jenis kelamin laki-laki dengan status pernikahan belum menikah masing-masing adalah 234 dan 465 sedangkan untuk jumlah angkatan kerja jenis kelamin perempuan dengan status pernikahan menikah masing-masing adalah 303 dan 393. Gambar 4.3 berikut ini menampilkan rangkuman jumlah angkatan kerja yang bekerja dan tidak bekerja berdasarkan jenis kelamin dan pengalaman pelatihan kerja, yaitu:

		Jenis Kelamin		0	1
		Laki-Laki	Perempuan		
Pengalaman Pelatihan Kerja	Pernah	71	99	79	91
	Tidak Pernah	625	600	617	608
	Grand Total	696	699	696	1395

Gambar 4.3 Data Cube berdasarkan Jenis Kelamin dan Pengalaman Pelatihan Kerja

Berdasarkan Gambar 4.3 di atas diperoleh bahwa jumlah angkatan kerja yang bekerja dengan jenis kelamin laki-laki dan perempuan masing-masing adalah 456 dan 240 sedangkan jumlah angkatan kerja yang tidak bekerja masing-masing adalah 456 dan 243. Untuk jumlah angkatan kerja yang bekerja dengan pernah dan tidak pernah mengikuti pengalaman pelatihan kerja masing-masing adalah 99 dan 600 sedangkan jumlah angkatan kerja yang tidak bekerja masing-masing adalah 71 dan 625. Untuk jumlah angkatan kerja jenis kelamin laki-laki dengan pernah dan tidak pernah mengikuti pengalaman pelatihan kerja masing-masing adalah 91 dan 608 sedangkan untuk jumlah angkatan kerja jenis kelamin perempuan dengan pernah dan tidak pernah mengikuti pengalaman pelatihan kerja masing-masing adalah 79 dan 617.

### Tahapan Analisis Data

Analisis data yang digunakan pada penelitian ini adalah Metode CART (*Classification And Regression Tree*) dengan langkah-langkah sebagai berikut:

### Membagi data menjadi dua bagian yaitu data learning dan data data testing

Pembagian data *learning* dan data *testing* pada penelitian ini dapat dilihat pada Tabel 4.5 sebagai berikut:



Tabel 4.5 Perbandingan data *learning* dan data *testing*

No	Perbandingan		Akurasi
	Data <i>Learning</i>	Data <i>Testing</i>	
1	95%	5%	29,87%
2	90%	10%	69,68%
3	85%	15%	66,37%
4	80%	20%	61,93%
5	75%	25%	63,82%

Berdasarkan pada Tabel 4.2 di atas perbandingan data testing dan data learning di atas diperoleh proporsi data learning dan data testing yang memiliki nilai akurasi tertinggi adalah perbandingan 90%:10%, sehingga dalam penelitian ini akan digunakan proporsi data learning dan data testing adalah 90%:10%. Dimana data *learning* sebanyak 1.395 sampel dan data *testing* sebanyak 155 sampel.

1. Membentuk pohon klasifikasi maksimal dengan langkah-langkah sebagai berikut:
  - a. Melakukan pemilihan pemilah berdasarkan aturan pemilihan indeks gini yang selanjutnya dievaluasi menggunakan kriteria *goodness of split*.

Pada tahapan ini akan dilakukan pemilihan pemilah dengan menggunakan indeks gini dengan calon pemilah dapat dilihat pada Tabel 4.6 sebagai berikut:

Tabel 4.6 Data calon simpul kiri dan simpul kanan

No.	Calon simpul kiri	Calon simpul kanan
1	Jenis kelamin=Laki-laki	Jenis kelamin=Perempuan
2	Tingkat pendidikan= Rendah ( $\leq SD$ )	Tingkat pendidikan = Sedang ( $SMP - SMA$ ), Tinggi (Sarjana)
3	Tingkat pendidikan= Sedang ( $SMP - SMA$ )	Tingkat pendidikan = Rendah ( $\leq SD$ ), Tinggi (Sarjana)
4	Tingkat pendidikan= Tinggi (Sarjana)	Tingkat pendidikan = Rendah ( $\leq SD$ ), Sedang ( $SMP - SMA$ )
5	Status pernikahan= Belum Menikah	Status pernikahan= Menikah
6	Pengalaman pelatihan kerja = Pernah	Pengalaman pelatihan kerja = Tidak pernah

Berdasarkan Tabel 4.6 di atas dapat dijelaskan bahwa calon simpul kiri adalah jenis kelamin laki-laki, tingkat pendidikan rendah, tingkat pendidikan sedang, tingkat pendidikan tinggi, status pernikahan belum menikah, dan pengalaman pelatihan kerja pernah. Sedangkan untuk calon simpul kanan, jenis kelamin perempuan, tingkat pendidikan sedang dan tinggi, tingkat pendidikan rendah dan tinggi, tingkat pendidikan rendah dan sedang, status

pernikahan menikah, dan pengalaman pelatihan kerja tidak pernah. Dari calon simpul diatas, akan ditentukan simpul mana yang akan menjadi parent node pada pohon keputusan yang terbentuk.

### Gain Information

Sebelum mencari nilai *gain information*, untuk memudahkan sebaiknya terlebih dahulu mencari nilai probabilitas untuk setiap simpul dengan menggunakan persamaan (2.4) dan (2.5) seperti pada simpul pertama diperoleh nilai sebagai berikut:

$$P_L = \frac{\text{calon simpul kiri}}{\text{data latihan}} = \frac{696}{1395} = 0,49892$$

$$P_R = \frac{\text{calon simpul kanan}}{\text{data latihan}} = \frac{699}{1395} = 0,50108$$

$$P(j|t_L) = \frac{n_j(t)}{n(t)} = \frac{456}{696} = 0,65517$$

$$P(j|t_R) = \frac{n_i(t)}{n(t)} = \frac{243}{699} = 0,34764$$

Dengan cara yang sama nilai untuk simpul yang lain adalah disajikan pada Tabel 4.7 sebagai berikut:

Tabel 4.7 Perhitungan probabilitas simpul

Simpul	$P_L$	$P_R$	Kelas	$P(j t_L)$	$P(j t_R)$
1	$\frac{696}{1395} = 0,49892$	$\frac{699}{1395} = 0,50108$	Bekerja	$\frac{456}{696} = 0,65517$	$\frac{243}{699} = 0,34764$
			Tidak Bekerja	$\frac{240}{696} = 0,34483$	$\frac{456}{699} = 0,65236$
2	$\frac{319}{1395} = 0,22867$	$\frac{1076}{1395} = 0,77133$	Bekerja	$\frac{172}{319} = 0,53918$	$\frac{527}{1076} = 0,48978$
			Tidak Bekerja	$\frac{147}{319} = 0,46082$	$\frac{549}{1076} = 0,51022$
3	$\frac{790}{1395} = 0,56631$	$\frac{605}{1395} = 0,43369$	Bekerja	$\frac{350}{790} = 0,44304$	$\frac{340}{605} = 0,56198$
			Tidak Bekerja	$\frac{440}{790} = 0,55696$	$\frac{265}{605} = 0,43802$
4	$\frac{286}{1395} = 0,20502$	$\frac{1109}{1395} = 0,79498$	Bekerja	$\frac{177}{286} = 0,61888$	$\frac{522}{1109} = 0,47069$
			Tidak Bekerja	$\frac{109}{286} = 0,38112$	$\frac{587}{1109} = 0,52931$
5	$\frac{537}{1395} = 0,38493$	$\frac{858}{1395} = 0,61507$	Bekerja	$\frac{213}{537} = 0,39665$	$\frac{466}{858} = 0,54312$
			Tidak Bekerja	$\frac{324}{537} = 0,60335$	$\frac{392}{858} = 0,45688$
6	$\frac{170}{1395} = 0,1219$	$\frac{1225}{1395} = 0,8781$	Bekerja	$\frac{99}{170} = 0,58235$	$\frac{600}{1225} = 0,49019$
			Tidak Bekerja	$\frac{71}{170} = 0,41765$	$\frac{625}{1225} = 0,50981$

Dengan menggunakan nilai pada Tabel 4.7 Selanjutnya untuk nilai *gain information* yang digunakan untuk mendeteksi atribut-atribut yang paling banyak memiliki informasi berdasarkan kelas tertentu. Untuk mencari nilai *gain information* untuk setiap atribut menggunakan rumus pada persamaan (2.1) sebagai berikut:

$$GI(jk_L) = - \sum_{i=1}^m P(j|t_L) \log_2 P(j|t_L) \\ = - \{ (0,65517) \log_2 (0,65517) + (0,34483) \log_2 (0,34483) \} \\ = 0,99979$$

Perhitungan nilai *gain information* untuk simpul yang lain dilakukan dengan cara yang sama sehingga diperoleh nilai seperti pada Tabel 4.8 sebagai berikut:

Tabel 4.8 Perhitungan *gain information*

Simpul	$P_L$	$P_R$	Kelas	$GI(P_L)$	$GI(P_R)$
1	Laki-laki	Perempuan	Bekerja	0,99979	0,99881
			Tidak Bekerja		
2	Tingkat pendidikan rendah	Tingkat pendidikan sedang dan tinggi	Bekerja	0,99555	0,99971
			Tidak Bekerja		
3	Tingkat pendidikan sedang	Tingkat pendidikan rendah dan tinggi	Bekerja	0,99062	0,98289
			Tidak Bekerja		
4	Tingkat pendidikan tinggi	Tingkat pendidikan rendah dan sedang	Bekerja	0,95882	0,99752
			Tidak Bekerja		
5	Status pernikahan belum menikah	Status pernikahan menikah	Bekerja	0,96896	0,98723
			Tidak Bekerja		
6	Pengalaman pelatihan kerja pernah	Pengalaman pelatihan kerja tidak pernah	Bekerja	0,94035	0,93971
			Tidak Bekerja		

Berdasarkan Tabel 4.8 di atas diperoleh bahwa nilai *gain information* dapat dilihat bahwa atribut yang paling banyak memberikan informasi adalah atribut jenis kelamin laki-laki yaitu sebesar 0,99979 begitupun atribut yang lain juga memiliki nilai yang tidak berbeda jauh dari nilai tersebut.

### Indeks Gini

Dengan menggunakan nilai pada Tabel 4.7, selanjutnya mencari nilai indeks gini untuk setiap simpul dengan menggunakan persamaan (2.2) sehingga diperoleh nilai indeks gini untuk calon simpul pertama adalah:

$$i(t) = 1 - \sum_{j=1}^2 P^2(j|t)$$

$$= 1 - (0,49892)^2 - (0,50108)^2 = 0,49999$$

Perhitungan indeks gini tersebut dilakukan untuk semua calon simpul sehingga diperoleh nilai yang dapat dilihat pada Tabel 4.7 sebagai berikut:

Tabel 4.9 Perhitungan Indeks Gini

Simpul	$P_L$	$P_R$	$i(t)$
1	Laki-laki	Perempuan	0,49999
2	Tingkat pendidikan rendah	Tingkat pendidikan sedang dan tinggi	0,35276
3	Tingkat pendidikan sedang	Tingkat pendidikan rendah dan tinggi	0,49121
4	Tingkat pendidikan tinggi	Tingkat pendidikan rendah dan sedang	0,32597
5	Status pernikahan belum menikah	Status pernikahan menikah	0,47353
6	Pengalaman pelatihan kerja pernah	Pengalaman pelatihan kerja tidak pernah	0,21418

### Goodness of Split

Selanjutnya dilakukan pemilihan pemilah atau calon simpul yang akan menjadi parent node atau simpul akar dengan menggunakan kriteria *goodness of split*. Untuk menghitung nilai *goodness of split* calon simpul pertama maka menggunakan persamaan (2.3) maka diperoleh :

$$i(t_L) = 1 - (0,65517)^2 - (0,34483)^2 \\ = 0,45184$$

$$i(t_R) = 1 - (0,34764)^2 - (0,65236)^2 \\ = 0,45357$$

$$\phi(s,t) = i(t) - P_L i(t_L) - P_R i(t_R) \\ = 0,49999 - 0,22543 - 0,22727 \\ = 0,04729$$

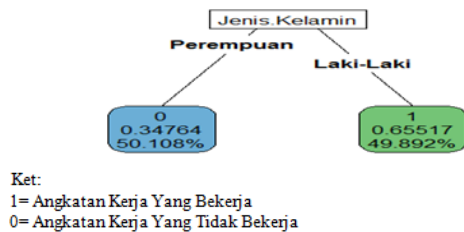
Sehingga diperoleh nilai *goodness of split* untuk setiap calon simpul dapat dilihat pada Tabel 4.10 sebagai berikut:

Tabel 4.10 *Goodness of split*

Simpul	$\phi(s,t)$	Kriteria kebaikan
1	0,04729	1
2	-0,14637	4
3	0,00001	2
4	-0,16687	5
5	-0,01281	3
6	-0,28366	6

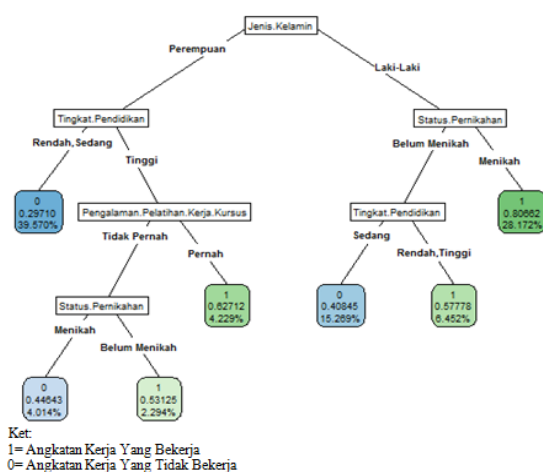
Berdasarkan hasil perhitungan nilai *goodness of split* pada Tabel 4.10 di atas untuk setiap calon

simpul diperoleh bahwa calon simpul yang tertinggi adalah calon simpul 2 sebesar **0,04729**, sehingga calon simpul ke 2 yaitu jenis kelamin ini akan menjadi parent node. Simpul ke 2 ini akan bercabang menjadi cabang kiri yaitu jenis kelamin laki-laki dan calon cabang kanan yaitu jenis kelamin perempuan dapat dilihat pada Gambar 4.4 berikut:



Gambar 4.4 Proses pemecahan pada parent node

Berdasarkan pada Gambar 4.4 di atas diperoleh bahwa parent node jenis kelamin yang terdiri dari node 2 yaitu perempuan dan node 3 yaitu laki-laki. Dimana dapat dilihat pada gambar 4.4 diatas bahwa nilai probabilitas untuk angkatan kerja perempuan yang bekerja sebanyak 0,34764 dari total angkatan kerja perempuan sebanyak 50,108% sedangkan untuk angkatan kerja laki-laki yang bekerja sebanyak 0,65517 dari total angkatan kerja laki-laki sebanyak 49,892% seperti yang ditunjukkan pada Tabel 4.7. Proses pemilihan akan terus berlanjut secara rekursif dengan cara yang sama pada node atau simpul yang lain sampai pada simpul akhir atau terminal node. Setelah dilakukan proses rekursif terhadap node-node yang lain dan sampai pada terminal node maka diperoleh pohon keputusan dengan 7 terminal node yang dapat dilihat pada Gambar 4.5 berikut:



Gambar 4.5 Struktur Pohon Keputusan

Berdasarkan struktur pohon pada Gambar 4.5 di

atas diperoleh bahwa parent node atau node 1 adalah jenis kelamin yang terdiri dari perempuan dan laki-laki yang masing-masing ditunjukkan pada node 2 dan node 3. Setelah mengalami proses pemecahan maka node 2 bercabang menjadi node 4 tingkat pendidikan (rendah dan sedang) yang menjadi terminal node dan node 5 tingkat pendidikan tinggi. Begitupula pada node 3 bercabang menjadi node 6 status pernikahan belum menikah dan node 7 status pernikahan menikah yang menjadi terminal node. Kemudian pada node 5 bercabang menjadi node 8 pengalaman pelatihan kerja (tidak pernah) dan node 9 .pengalaman pelatihan kerja (pernah) yang menjadi terminal node. Pada node 6 bercabang menjadi node 10 tingkat pendidikan sedang dan node 11 tingkat pendidikan rendah dan tinggi yang kedua node ini menjadi terminal node. Pada node 8 bercabang menjadi node 8 pengalaman pelatihan kerja (tidak pernah) bercabang menjadi node 12 status pernikahan menikah dan node 13 status pernikahan belum menikah yang kedua node ini menjadi terminal node. Struktur pohon pada Gambar 4.5 dapat dituliskan kedalam tabel sebagai berikut:

Tabel 4.11 Tabel node pada struktur pohon keputusan

Node	Nama	Keterangan
1	Jenis kelamin	Non terminal node
2	Jenis kelamin perempuan	Non terminal node
3	Jenis kelamin laki-laki	Non terminal node
4	Tingkat pendidikan rendah dan sedang	Terminal node
5	Tingkat pendidikan tinggi	Non terminal node
6	Status pernikahan belum menikah	Non terminal node
7	Status pernikahan menikah	Terminal node
8	Pengalaman pelatihan kerja tidak pernah	Non terminal node
9	Pengalaman pelatihan kerja pernah	Terminal node
10	Tingkat pendidikan sedang	Terminal node
11	Tingkat pendidikan rendah dan tinggi	Terminal node
12	Status pernikahan menikah	Terminal node
13	Status pernikahan belum menikah	Terminal node

Berdasarkan pada Tabel 4.11 di atas dapat dijelaskan bahwa simpul terminal atau terminal node merupakan simpul terakhir yang tidak terdapat penurunan keheterogenan secara berarti berdasarkan kriteria *goodness of split*. Pada pohon keputusan pada Gambar 4.5 diatas dapat dilihat bahwa simpul yang menjadi non terminal node adalah node 1,2,3,5,6, dan 8 sedangkan

yang menjadi terminal node adalah node 4,9,10,11,12,dan 13.

### Pelabelan Kelas

Pelabelan kelas pada terminal node berdasarkan aturan jumlah terbanyak dari tiap kelas yang ada pada variabel respon. Penandaan label kelas berguna untuk mengidentifikasi tiap node pada kelas tertentu. Penandaan label kelas didasarkan pada persamaan (2.7) dan dapat dilihat pada Tabel 4.2 yang menunjukkan bahwa pada jumlah pengamatan terbanyak pada simpul akar atau root node (jenis kelamin) adalah angkatan kerja yang bekerja yang diberi label dengan angka 1 begitu pula dengan node-node yang lain. Berdasarkan pada Gambar 4.5 struktur pohon keputusan maka selanjutnya akan disajikan penandaan label kelas untuk setiap node khususnya penandaan label kelas pada terminal node yang dapat dilihat pada Tabel 4.12 berikut:

Tabel 4.12 Penandaan Label Kelas

Node	Simpul	Label Kelas
1	Jenis kelamin	1
2	Jenis kelamin perempuan	0
3	Jenis kelamin laki-laki	1
4	Tingkat pendidikan rendah dan sedang	0
5	Tingkat pendidikan tinggi	1
6	Status pernikahan belum menikah	0
7	Status pernikahan menikah	1
10	Pengalaman pelatihan kerja tidak pernah	0
11	Pengalaman pelatihan kerja pernah	1
12	Tingkat pendidikan sedang	0
13	Tingkat pendidikan rendah dan tinggi	1
14	Status pernikahan menikah	0
15	Status pernikahan belum menikah	1

Berdasarkan pada Tabel 4.12 di atas dapat dijelaskan bahwa node dengan label kelas angkatan kerja yang bekerja berturut-turut adalah node 1,3,5,7,9,11 dan 13. Sedangkan untuk node dengan label kelas angkatan kerja yang tidak bekerja berturut-turut adalah node 2,4,6,8,10 dan 12.

### Pemangkasan pohon klasifikasi berdasarkan kriteria cost complexity minimum.

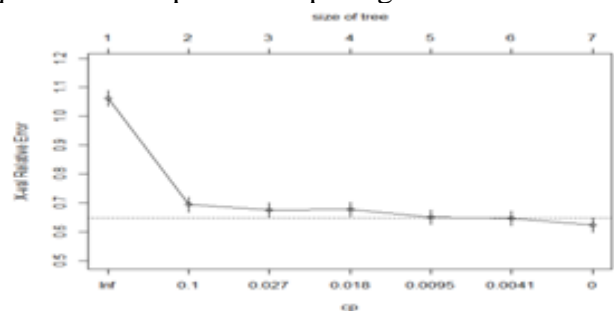
Dalam mendapatkan pohon yang baik maka dapat dilakukan dengan melakukan pemangkasan pohon berdasarkan persamaan (2.10) cabang yang akan dilakukan pemangkasan adalah cabang yang memiliki nilai *complexity parameter minimum*. Dengan menggunakan

persamaan (2.9) dan (2.10) maka diperoleh nilai *complexity parameter* pohon T disajikan pada Tabel 4.13 sebagai berikut:

Tabel 4.13 Nilai *complexity parameter* pohon T

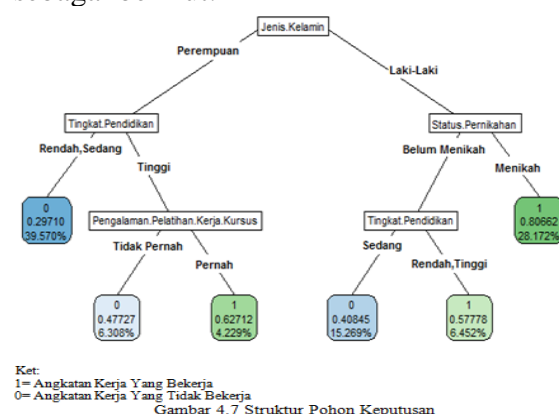
No.	<i>Complexity parameter</i>
1	0,3060345
2	0,0359195
3	0,0201149
4	0,0158046
5	0,0057471
6	0,0028736

Berdasarkan pada Tabel 4.13 di atas diperoleh nilai *complexity parameter* untuk setiap split berturut-turut adalah 0,3060345; 0,0359195; 0,0201149; 0,0158046; 0,0057471; 0,0028736. Sehingga nilai *complexity parameter minimum* adalah 0,0028736. plot nilai *complexity parameter* dapat dilihat pada gambar berikut:



Gambar 4.6 Plot *complexity parameter*

Berdasarkan pada Gambar 4.6 plot *complexity parameter* di atas maka pohon yang akan dipangkas adalah subpohon yang memiliki nilai *complexity parameter* yang ditunjukkan pada gambar dimana ditunjukkan bahwa nilai *complexity minimum* yang akan dipangkas adalah 0,0028736 yaitu nilai yang kurang dari 0,0057471. Sehingga setelah dilakukan pemangkasan maka diperoleh struktur pohon sebagai berikut:



Gambar 4.7 Struktur Pohon Keputusan

Berdasarkan struktur pohon pada Gambar 4.7 di



atas diperoleh bahwa parent node atau node 1 adalah jenis kelamin yang terdiri dari perempuan dan laki-laki yang masing-masing ditunjukkan pada node 2 dan node 3. Setelah mengalami proses pemecahan maka node 2 bercabang menjadi node 4 tingkat pendidikan (rendah dan sedang) yang menjadi terminal node dan node 5 tingkat pendidikan tinggi. begitupula pada node 3 bercabang menjadi node 6 status pernikahan belum menikah dan node 7 status pernikahan menikah yang menjadi terminal node. Kemudian pada node 5 bercabang menjadi node 8 pengalaman pelatihan kerja (tidak pernah) dan node 9 .pengalaman pelatihan kerja (pernah) yang kedua node ini menjadi terminal node. Pada node 6 bercabang menjadi node 10 tingkat pendidikan sedang dan node 11 tingkat pendidikan rendah dan tinggi yang kedua node ini menjadi terminal node.

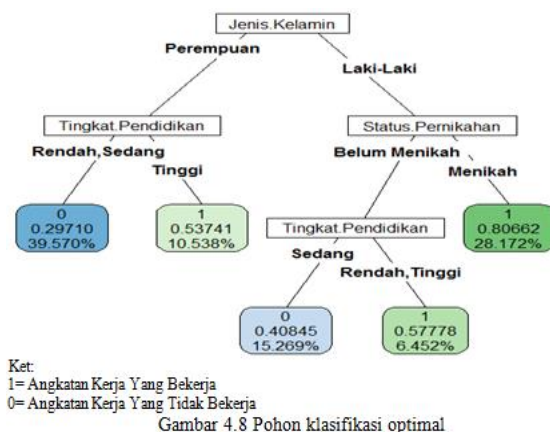
#### Memilih pohon klasifikasi optimal.

Setelah dilakukan pemangkasan pohon maka akan dilakukan pemilihan pohon optimal dengan menggunakan penduga validasi silang lipat V (*cross-validation V-fold estimate*) karena ukuran data yang kecil yang kurang dari 3000 data. Dengan menggunakan persamaan pada (2.12) dan (2.13) maka diperoleh nilai yang disajikan pada Tabel 4.14 sebagai berikut:

Tabel 4.14 Tabel Proporsi kesalahan *V-fold Estimate*

V-fold	Proporsi kesalahan klasifikasi
1	0,3297491
2	0,311828
3	0,2831541
4	0,3548387
5	0,3297491

Berdasarkan pada Tabel 4.14 di atas dapat dilihat bahwa nilai proporsi kesalahan untuk 5-fold berturut-turut adalah 0,3297491; 0,311828; 0,2831541; 0,3548387 dan 0,3297491. Sehingga diperoleh pohon optimal dengan nilai proporsi kesalahan 5-fold terkecil adalah 0,2831541. Sehingga pohon klasifikasi optimal dapat digambarkan sebagai berikut:



Berdasarkan struktur pohon pada Gambar 4.8 di atas diperoleh bahwa parent node atau node 1 adalah jenis kelamin yang terdiri dari perempuan dan laki-laki yang masing-masing ditunjukkan pada node 2 dan node 3. Setelah mengalami proses pemecahan maka node 2 bercabang menjadi node 4 tingkat pendidikan (rendah dan sedang) dan node 5 tingkat pendidikan tinggi yang kedua node ini menjadi terminal node. begitupula pada node 3 bercabang menjadi node 6 status pernikahan belum menikah dan node 7 status pernikahan menikah yang menjadi terminal node. Pada node 6 bercabang menjadi node 10 tingkat pendidikan sedang dan node 11 tingkat pendidikan rendah dan tinggi yang kedua node ini menjadi terminal node.

Untuk cabang jenis kelamin perempuan yang memiliki tingkat pendidikan (rendah dan sedang), dan tinggi pada Gambar 4.8 di atas dapat ditampilkan pada Tabel 4.15 sebagai berikut:

Tabel 4.15 Jumlah angkatan kerja berdasarkan cabang jenis kelamin dan tingkat pendidikan

Tingkat Pendidikan	Observasi		Bekerja		Tidak bekerja	
	N	%	N	%	N	%
Rendah dan Sedang	552	39,57%	164	29,71%	388	70,29%
Tinggi	147	10,53%	79	53,74%	68	46,26%

Berdasarkan pada Tabel 4.15 di atas dapat dijelaskan bahwa jumlah angkatan kerja jenis kelamin perempuan dengan tingkat pendidikan rendah dan sedang sebanyak 552 (39,57%, n=1395). Dimana dari jumlah tersebut angkatan kerja yang bekerja hanya 164 (26,57%, n=552) sedangkan angkatan kerja yang tidak bekerja sebanyak 388 (70,29%, n=552). Untuk angkatan kerja jenis kelamin perempuan dengan tingkat



pendidikan tinggi sebanyak 147 (10,53%, n=1395). Dimana dari jumlah tersebut angkatan kerja yang bekerja sebanyak 79 (53,74%, n=147) sedangkan angkatan kerja yang tidak bekerja hanya 68 (46,26%, n=147).

Untuk cabang jenis kelamin laki-laki yang memiliki status pernikahan menikah dan belum menikah pada Gambar 4.8 di atas dapat ditampilkan pada Tabel 4.16 sebagai berikut:

Tabel 4.16 Jumlah angkatan kerja berdasarkan cabang jenis kelamin dan status pernikahan

Status Pernikahan	Observasi		Bekerja		Tidak bekerja	
	N	%	N	%	N	%
Menikah	393	28,17%	317	80,66%	76	19,34%
Belum Menikah	303	21,72%	139	45,87%	164	54,13%

Berdasarkan pada Tabel 4.15 di atas dapat dijelaskan bahwa jumlah angkatan kerja jenis kelamin laki-laki dengan status pernikahan menikah sebanyak 393 (28,17%, n=1395). Dimana dari jumlah tersebut angkatan kerja yang bekerja sebanyak 317 (80,66%, n=393) sedangkan angkatan kerja yang tidak bekerja hanya 76 (19,34%, n=393). Untuk angkatan kerja jenis kelamin laki-laki dengan status pernikahan belum menikah sebanyak 303 (21,72%, n=1395). Dimana dari jumlah tersebut angkatan kerja yang bekerja hanya 139 (45,87%, n=303) sedangkan angkatan kerja yang tidak bekerja sebanyak 164 (54,13%, n=303).

Untuk cabang pada jenis kelamin laki-laki dengan status pernikahan menikah yang memiliki tingkat pendidikan sedang, rendah dan tinggi pada Gambar 4.8 di atas dapat ditampilkan pada Tabel 4.16 sebagai berikut:

Tabel 4.17 Jumlah angkatan kerja berdasarkan cabang jenis kelamin, status pernikahan dan tingkat pendidikan

Tingkat Pendidikan	Observasi		Bekerja		Tidak bekerja	
	N	%	N	%	N	%
Sedang	213	15,27%	87	40,84%	126	59,16%
Rendah dan Tinggi	90	6,45%	52	57,77%	38	42,23%

Berdasarkan pada Tabel 4.15 di atas dapat dijelaskan bahwa jumlah angkatan kerja jenis kelamin laki-laki, status pernikahan belum menikah dengan tingkat pendidikan sedang sebanyak 213 (15,27%, n=1395). Dimana dari

jumlah tersebut angkatan kerja yang bekerja hanya 87 (40,84%, n=213) sedangkan angkatan kerja yang tidak bekerja sebanyak 126 (59,16%, n=213). Untuk angkatan kerja jenis kelamin laki-laki, status pernikahan belum menikah dengan tingkat pendidikan rendah dan tinggi sebanyak 90 (6,45%, n=1395). Dimana dari jumlah tersebut angkatan kerja yang bekerja sebanyak 52 (57,77%, n=90) sedangkan angkatan kerja yang tidak bekerja hanya 38 (42,23%, n=90).

2. Menghitung ketepatan klasifikasi pohon dengan menggunakan *sensitivity*, *specificity*, dan akurasi.

Ketepatan klasifikasi untuk data prediksi ditunjukkan pada Tabel 4.15 berikut:

Tabel 4.18 Ketepatan klasifikasi untuk data prediksi

Observasi	Prediksi		Total
	Angkatan kerja yang tidak bekerja	Angkatan kerja yang bekerja	
Angkatan kerja yang tidak bekerja	48	29	77
Angkatan kerja yang bekerja	18	60	78
Total	66	89	155

Berdasarkan pada Tabel 4.18 di atas dapat dijelaskan bahwa untuk angkatan kerja yang bekerja berdasarkan prediksi sebesar 60 orang sedangkan untuk angkatan kerja yang tidak bekerja berdasarkan prediksi diperoleh 48 orang. Dengan menggunakan persamaan (2.16) untuk perhitungan akurasi, persamaan (2.14) untuk perhitungan *sensitivity*, dan persamaan (2.15) untuk perhitungan *specificity* pada data testing maka diperoleh hasil sebagai berikut:

$$Akurasi = \frac{48 + 60}{155} \times 100\% = 69,68\%$$

$$Sensitivity = \frac{48}{77} \times 100\% = 62,34\%$$

$$Specificity = \frac{60}{78} \times 100\% = 76,92\%$$

Dari perhitungan yang dilakukan, nilai akurasi yang digunakan untuk mengukur tingkat ketepatan klasifikasi untuk data prediksi sebesar 69,68% maka dapat dikatakan bahwa pohon optimal yang terbentuk mampu mengklasifikasikan data baru sebesar 69,68%.

Sedangkan untuk nilai *sensitivity* yang digunakan untuk mengukur ketepatan klasifikasi pada kelas angkatan kerja yang tidak bekerja sebesar 62,34% dan *specifity* yang digunakan untuk mengukur ketepatan klasifikasi pada kelas angkatan kerja yang bekerja sebesar 76,92%.

3. Menentukan jumlah angkatan kerja yang bekerja dan tidak bekerja berdasarkan jenis kelamin, tingkat pendidikan, status pernikahan dan pengalaman pelatihan kerja Berdasarkan pohon klasifikasi yang diperoleh pada Gambar 4.8 diperoleh jumlah angkatan kerja berdasarkan kelompok-kelompok berikut:

- a. Kelompok pertama yaitu angkatan kerja usia produktif dengan jenis kelamin perempuan dengan tingkat pendidikan (rendah dan sedang) sebanyak 552 (40,14%; n=1395), dengan jumlah angkatan kerja yang bekerja sebanyak 164 (26,57%, n=552) sedangkan jumlah angkatan kerja yang tidak bekerja sebanyak 388 (70,29%, n=552).
- b. Kelompok kedua yaitu angkatan kerja usia produktif dengan jenis kelamin perempuan dengan tingkat pendidikan tinggi sebanyak 147 (10,53%, n=1395), dengan jumlah tersebut angkatan kerja yang bekerja sebanyak 79 (53,74%, n=147) sedangkan jumlah angkatan kerja yang tidak bekerja sebanyak 68 (46,26%, n=147).
- c. Kelompok ketiga yaitu angkatan kerja usia produktif dengan jenis kelamin laki-laki dengan status pernikahan belum menikah dan tingkat pendidikan sedang sebanyak 213 (15,27%, n=1395), dengan jumlah angkatan kerja yang bekerja sebanyak 87 (40,84%, n=213) sedangkan jumlah angkatan kerja yang tidak bekerja sebanyak 126 (59,16%, n=213).
- d. Kelompok keempat yaitu angkatan kerja usia produktif dengan jenis kelamin laki-laki dengan status pernikahan belum menikah dan tingkat pendidikan rendah dan tinggi sebanyak 90 (6,45%, n=1395), dengan jumlah angkatan kerja yang bekerja sebanyak 52 (57,77%, n=90) sedangkan angkatan kerja yang tidak bekerja sebanyak 38 (42,23%, n=90).

- e. Kelompok kelima yaitu angkatan kerja usia produktif dengan jenis kelamin laki-laki dengan status pernikahan belum menikah sebanyak 393 (28,17%, n=1395), dengan jumlah angkatan kerja yang bekerja sebanyak 317 (80,66%, n=393) sedangkan angkatan kerja yang tidak bekerja sebanyak 76 (19,34%, n=393).

## 5. KESIMPULAN

Adapun kesimpulan pada penelitian ini adalah bahwa jumlah angkatan kerja di kota Makassar berdasarkan jenis kelamin, tingkat pendidikan, status pernikahan dan pengalaman pelatihan kerja dikelompokkan berdasarkan kelompok berikut :

- a. Kelompok pertama yaitu angkatan kerja usia produktif dengan jenis kelamin perempuan dengan tingkat pendidikan (rendah dan sedang) sebanyak 552 (40,14%; n=1395), dengan jumlah angkatan kerja yang bekerja sebanyak 164 (26,57%, n=552) sedangkan jumlah angkatan kerja yang tidak bekerja sebanyak 388 (70,29%, n=552).
- b. Kelompok kedua yaitu angkatan kerja usia produktif dengan jenis kelamin perempuan dengan tingkat pendidikan tinggi sebanyak 147 (10,53%, n=1395), dengan jumlah tersebut angkatan kerja yang bekerja sebanyak 79 (53,74%, n=147) sedangkan jumlah angkatan kerja yang tidak bekerja sebanyak 68 (46,26%, n=147).
- c. Kelompok ketiga yaitu angkatan kerja usia produktif dengan jenis kelamin laki-laki dengan status pernikahan belum menikah dan tingkat pendidikan sedang sebanyak 213 (15,27%, n=1395), dengan jumlah angkatan kerja yang bekerja sebanyak 87 (40,84%, n=213) sedangkan jumlah angkatan kerja yang tidak bekerja sebanyak 126 (59,16%, n=213).
- d. Kelompok keempat yaitu angkatan kerja usia produktif dengan jenis kelamin laki-laki dengan status pernikahan belum menikah dan tingkat pendidikan rendah dan tinggi sebanyak 90 (6,45%, n=1395), dengan jumlah angkatan kerja yang bekerja sebanyak 52 (57,77%, n=90) sedangkan

angkatan kerja yang tidak bekerja sebanyak 38 (42,23%, n=90).

- e. Kelompok kelima yaitu angkatan kerja usia produktif dengan jenis kelamin laki-laki dengan status pernikahan belum menikah sebanyak 393 (28,17%, n=1395), dengan jumlah angkatan kerja yang bekerja sebanyak 317 (80,66%, n=393) sedangkan angkatan kerja yang tidak bekerja sebanyak 76 (19,34%, n=393).

## 6. DAFTAR PUSTAKA

- [1] Abdul Kadir, Mohamad, "Perbandingan Performasi Algoritma Decision Tree CART dan CHAID", (Bandung: Institute Teknologi Bandung, Program Studi Teknik Informatika)
- [2] Aprilia, Tiara, dkk, 2015, "Klasifikasi ketepatan masa studi mahasiswa FMIPA Unpad angkatan 2001-2006 dengan menggunakan metode Classification and Regression Tree (CART)". (vol.11, No.2) hal. 10
- [3] Badan Pusat Statistik. 2018. "Kota Makassar Dalam Angka 2018"
- [4] Eka Pratiwi, Febti & Ismaini zain, 2014, "Klasifikasi Pengangguran Terbuka Menggunakan CART (Classification and Regression Tree) di Provinsi Sulawesi Utara", (Vol.3, No.1)
- [5] Faisal, Muhammad, dkk, 2008, "Perbandingan Kinerja Metode Klasifikasi Chi-square Automatic Interaction Detection (CHAID) dengan Metode Klasifikasi Algoritma C4.5 pada Studi Kasus : Penderita Diabetes Melitus Tipe 2 Di Samarinda Tahun 2015" (Samarinda: Universitas Mulawarman, Program Studi Statistik FMIPA, 2017 vol.8 No.2)
- [6] Hartati, Alia, dkk, 2012, "Analisis CART (Classification and Regression Tree pada Faktor-faktor yang mempengaruhi Kepala Rumah Tangga di Jawa Timur Melakukan Urbanisasi" (vo.1, No.1)
- [7] Han Jiawei, dkk, "Data Mining Concepts and Techniques Third Edition" (USA: Elsevier, 2012)
- [8] Holis Sumartini, Siti & Santi Wulan Purnami, 2015, "Penggunaan Metode Classification and Regression Tress (CART) untuk Klasifikasi Rekurensi Pasien Kanker Serviks di RSUD Dr. Soetomo Surabaya", (Surabaya: Vol.4 No.2)
- [9] Janecek, Andreas, 2009, "Efficient Features Reduction and Classification Methods" (Wiem: Universitas Wien, 2009)
- [10] Kumara, Rendragraha & Catur Supriyanto. "Klasifikasi Data Mining Untuk Penerimaan Seleksi Calon Pegawai Negeri Sipil 2014 Menggunakan Algoritma Decision Tree C4.4" (Universitas Dian Nuswantoro Semarang: Program Studi Teknik Informatika Fakultas Ilmu Komputer)
- [11] Kurnia Damayanti, Laila, "Aplikasi algoritma CART untuk Mengklasifikasi Data Nasabah Asuransi Jiwa Bersama BUMIPUTERA 1912 Surakarta", (Surakarta : Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Sebelas Maret)
- [12] Komariah, Nurul, dkk, "Classification and Regression Tree (CART) analysis Pada Penderita Skizofrenia di RSJKO Soeprapto Daerah Bengkulu" (Universitas Bengkulu: jurusan matematika Fakultas MIPA)
- [13] Lemon Stephenie C., dkk, 2003, "Classification and Regression Tree Analysis in Public Health: Methodological Review and Comparison With Logistic Regression" (University of Massachusetts Medical School)
- [14] Lewis, R., 2000, "An Introduction to Classification And Regression Tree (CART) Analysis", (Annual Meeting of the society for Academic Emergency Medicine in San Fransisco, California, Departement Of Emergency Medicine, California)
- [15] Mardiani, 2012, "Penerapan Klasifikasi dengan Algoritma CART untuk Prediksi Kuliah Bagi Mahasiswa baru". (Palembang: STMIK MDP, jurusan Sistem Informasi)
- [16] Nafi'iyah Nur, 2015, "Algoritma CART dalam Penentuan Pohon Keputusan

- Sertifikasi Guru*”(Universitas Lamongan: Teknik Informatika, Vol.7, No.2)
- [17] Pratiwi, Nikhen S.T, 2008, ” *Analisis Perbandingan Pohon Klasifikasi Antarametode Quest Dengan Metode Chaid*”(Bogor: Institut Pertanian Bogor, Departemen Statistika Fakultas MIPA)
- [19] Rokah, Lior & Oded Maimon, ”*Decision Trees*” (Tel-Aviv University :Department Of Industri Engineering”
- [20] Roswita Nazar, Reny, 2018 ”*Penerapan Metode CHAID (Chi-Squared Automatic Interaction Detection) dan CART (Classification and Regression Tree) Pada Kalsifikasi Preeklampsia*”(Yogyakarta: Universitas Islam Indonesia: Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam)
- [21] Tittonell, P., dkk, 2007, ”*Unravelling the effects of soil and crop management on maize productivity in smallholder agricultural systems of western Kenya—An application of classification and regression tree analysis*”( Afrika: Tropical Soil Biology and Fertility Institute of the International Centre for Tropical Agriculture (TSBF-CIAT))
- [22] Wathoni, Mahbubul, 2006, ”*Prediksi Kecenderungan Konsumen dalam Memilih Jenis Kendaraan (Roda Empat) Berdasarkan Spesifikasi Kendaraan Menggunakan Decision Tree dengan Metode Gini*”(Jakarta : Universitas Islam Negeri Syarif Hidayatullah Program studi Matematika Jurusan MIPA Fakultas Sains dan teknologi )
- [23] Yisehac, Yohannes & John Hodddinott, ”*Classification and Regression Tree An Introduction*”(USA: International Food Policy Research Institute, 1999)